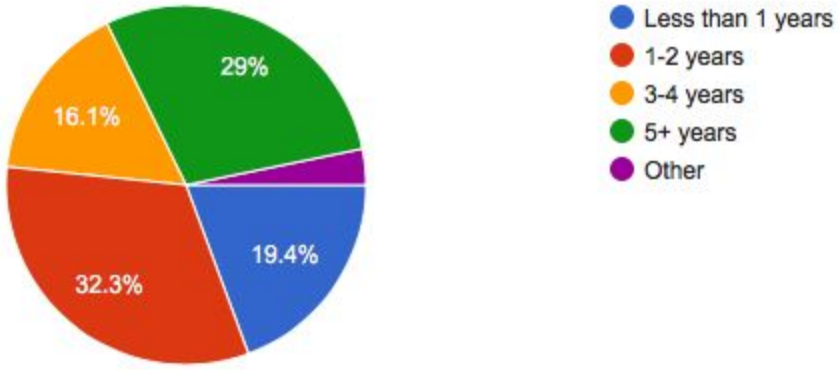
Community for Data Integration Feedback Session

Feedback collected September 14, 2016 during the CDI September Monthly Meeting
Contact: Leslie Hsu, lhsu@usgs.gov

What is CDI to you?

How long have you been a member of CDI?



How did you hear about CDI and decide to join?

- Personal invitation: 17
- Email list: 5
- Other (workshop, there at birth, don't remember): 7

What do you consider the main selling points of CDI?

The broad range of responses emphasized knowledge exchange and networking and collaboration. Some like the fact that it addresses official data management and standards at the Bureau level. Others stress the importance of interacting with like-minded peers informally.

- Awareness of USGS activities
- Comparing lessons learned, best practices
- Trainings
- Networking, Community aspect, Opportunity to collaborate
- Info on standards and data management
- Crosses all mission areas, efforts at cross-agency
- Interesting topics
- No commitments

What are your knowledge gaps for CDI?

- How non USGS folks can participate more fully
- What are the resources that CDI funded in the past, can I use them?
- Tighter links to the field
- Who are the active members, what mission areas are missing?
- What are the working groups up to?
- When do the working groups meet? I miss meetings as they fall off my radar.
- How to engage

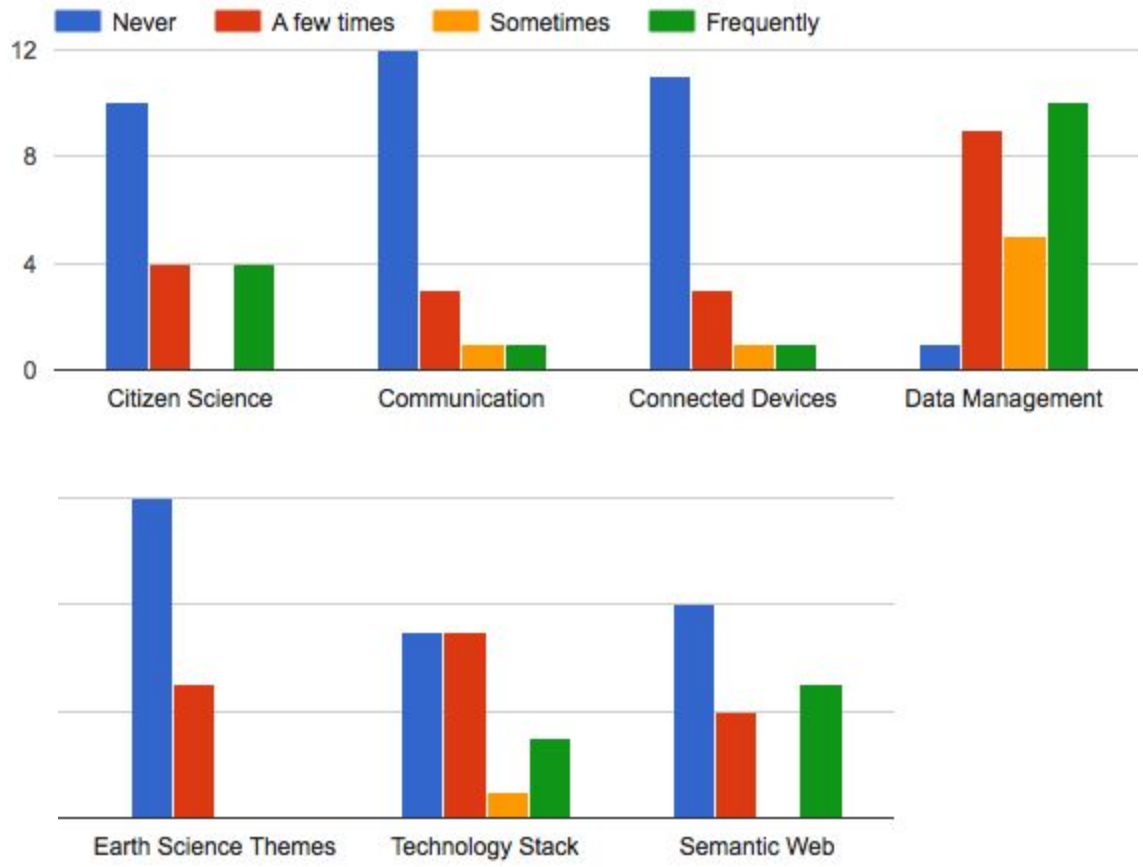
Misc suggestions

- What is the status of Integration of data sources, access, metadata, visualization among entities (other federal, universities)
- A Science Center tour. Visit each Center and make a pitch.
- Clean old content, make info accessible from a single location.
- Broadcast activities and achievements on USGS news blasts.
- Perhaps still too focused on "data integration' sense strict; we've outgrown the name.
- Need help with DMPs

Working Groups in FY17

Which Working Groups have you participated in?

These results are limited to the participants of the September Monthly call, but the most frequently attended working groups by this population appear to be Data Management and Semantic Web.



What activities would you like to see related to working groups at the FY17 Annual Meeting?

At the in-person FY17 Annual Meeting, respondents wanted face to face work sessions and time to discuss goals for the next year. The importance of working group presentations to the

entire group so that everyone understands all of the working groups was noted. Data release was a popular topic.

Other topics suggested

- Data Rescue
- Linked Data and Semantic technologies. How average researcher can leverage these.
- Internet of Things
- Communication Plan for CDI
- Scientific workflow
- HPC
- Use of cloud storage and compute resources (how-to)
- Data collection devices (wearables)
- Data Release - edge cases (images, video, summary data and graphs)
- Decision Support tools utilizing USGS datasets
- Open Innovation, crowdsourcing, citizen science, civic Hacking, prizes and challenges
- Citizen Science Data Quality.

Working group activity is reported at Monthly Meetings and on Blog posts. Do you feel there is enough information about working groups available to you?

- Yes: 9
- No: 5
- Other: 5 (Most of the “others” would appreciate more information.)

As a result of these responses, we will attempt to make the working group information more prominent in the upcoming monthly calls, and review the purpose of the working groups.

Other comments about working group activities you would like to see in FY17?

- Better summary of expertise or help needed in groups
- Do more work across working groups
- Give more information during the monthly meetings
- Want more information on data releases
- Coordinate with NSF Big Data Hubs
- More detailed working group reports for interesting things (deep dives)

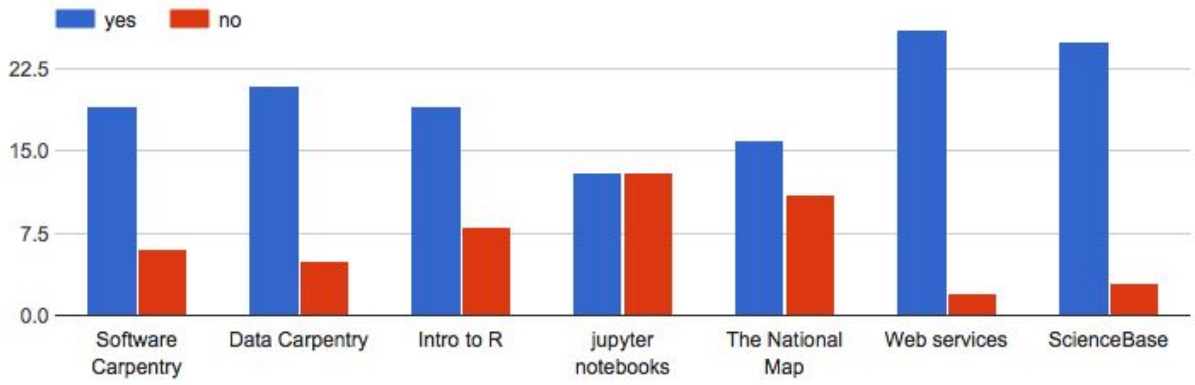
Trainings

What skills do you think belong in baseline knowledge for analysts and scientists, especially related to communication, web tools, and visualization?

Most people did not respond to this or said they didn't have time to think deeply for a good answer, but this might be a question to ask again at another time.

- Python, R, JS
- JSON, WebAPIs, RESTful Web
- AGOL

Are you interested in attending training on the following topics?



The participants on the call had greatest interest in learning about web services, ScienceBase, and Intro to R.

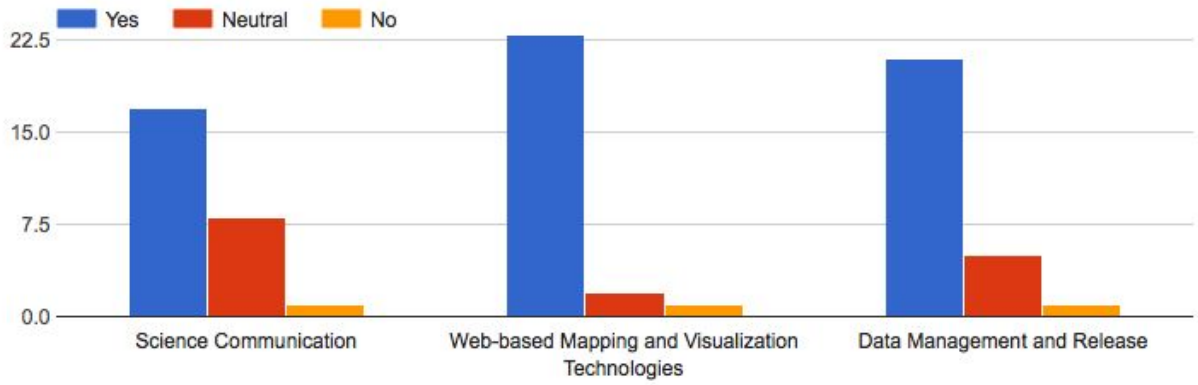
Comments

- I need to know more about what these topics actually are
- Suggested topic: "How to communicate within the USGS?"
- Personally I'm not really into training
- What are the enterprise tools we have available at our agency?
- Look at DataNet Federation tools: DataOne, CyVverse, iRODs.org - South Big Data Hub is organizing demos of these tools

- Engage with local meetups, hackathons, other open events, facilitate local events near our science centers.

Hands-on Sessions

Are you interested in the follow topics for a half-day hands-on sessions?



There was interest in all of these topics, but web-based mapping and visualization came out on top, and data management and release second.

More detail about Science Communication, Web-based Mapping and Visualization, and Data Management and Release

Science Communication

- What USGS resources can we use to be more effective at science communication?
- Session to practice developing stories about the work that we do.
- Science communication best practices

Web-based Mapping and Visualization

- Posting ArcGIS Online layers to the public
- Hosting recurring hackathons or open lab events to produce proof of concepts with USGS data
- Story Map demos
- How to start and maintain web development skills.

- Present currently used web-based mapping and viz tools that integrate data with metadata
- Demo open-source web-based mapping, how to incorporate these tools
- Basics of serving web-based vector data, learn enough to evaluate skills needed to accomplish tasks

Data Management and Release

- How to automate DM Release workflows
- Use of IPDS and ScienceBase to accomplish data release
- Review current status of tools available for data management and release
- Best practices for authoring metadata, standardized procedure for data release
- Is USGS developing standards for Data Management and Release?
- How are other centers (biology) handling the complexities of funding streams and data release obligations
- How to navigate the FSP process

Other:

- Coding topics.
- Writing modularized and testable code to process USGS data, version control.
- Field teams highlight their webtools that they use for their partners.
- Data Rescue - sharing info and equipment needed
- Metadata training
- Come up with ways for post-training activities to help participants continue to learn and grow.
- Project planning and data management.
- Workflows for mobile mapping using Arc GIS Online

Miscellaneous

- Working with Federal, State, local, tribal, and etc. partners
- “Doctor’s office” time to sit with people and “pick their brain”
- It’s important to build relationships across organizational boundaries.
- Synthesize the state of Data Integration, overview different data system differences and similarities
- With CDI work with the National Map to include other USGS thematic data - geology, geophysics, mineral resources, hazards...

Volunteered “lightning talks”

- Peter Schweitzer: questions that we get asked a lot but which for good reasons, we don’t have answers. Do others have similar experiences?
- Jake Weltzin: Delivery of real-time biodiversity info
- Sophia Liu: Visualizing Critical Minerals project
- Geoff Phelps: WorldWind Virtual Globe subsurface capability
- Jeremiah Lant: cookiecutter template for Python projects: https://github.com/usgs/cookiecutter-python-package, or Version control with Git or using a testing library for Python called pytest for testing your code

Other topics from the CDI-Theme Mapping Innovation Session

- Semantic Web
- Leaflet
- Shiny
- D3
- Map story
- Training workshops on Javascript/python/etc
- Expand the Scientist’s Challenge to a Data Challenge or Mapping Challenge
- Easy methods to make our data available to each other and our users
- Federating discovery of USGS data holdings
- Making reports and journal articles discoverable through JSTOR and Google Scolar
- Report out of all mapping innovation sessions
- Discussion for controlling costs of geospatial applications and services in the cloud - out-of-box ideas like serverless static vector tiles, caching techniques
- Establishing baseline programming knowledge for analysts and scientists
- Engaging external data science communities through hackathons